NOTES FOR THREE LANSDOWNE LECTURES ON MATRIX ALGEBRA

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

BU-987-M

Abstract

Notes used as hand-outs and transparencies for three lectures as Lansdowne Guest Speaker at University of Victoria, British Columbia, September, 1988:

I: Matrix Algebra: mathematics as useful as calculus

II: Some of today's popular matrix algebra.

III: Using matrix algebra in statistics.

Lecture 1

MATRIX ALGEBRA: MATHEMATICS AS USEFUL AS CALCULUS

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

Calculus — well recognized; great practical value

— so also with matrix algebra

— tremendously useful

— becoming ever more so — computers

Matrix algebra — an algebra

— deals with arrays of numbers

— as entities in their own right

<u>Matrix</u>: rectangular array of numbers

$$\mathbf{A} = \begin{bmatrix} 2 & -11 & 14^5 & -6.17892 & x^2-17x \\ 0 & t^3-y^2 & \sqrt{z} & 31 & e^t \end{bmatrix}.$$

<u>Order</u>: $A_{2\times 5}$ is "2 by 5": 2 rows, 5 columns.

Origins: Cayley, A. (1855) *Journal fur die neine und angewandte Mathematik* **50**, 282-285.
Cayley, A. (1858) *Philosophical Transactions of the Royal Society of London* **148**, 17-37.
Sylvester, J.J. (1850) *Philosophical Magazine* **37**,363-370.

A FEW (OF MANY) SPECIAL FORMS

Square: $\mathbf{B} = \begin{bmatrix} 0 & 1 & 3 \\ 9 & 6 & 9 \\ 2 & 4 & 9 \end{bmatrix}$.

Row vector: $\mathbf{x}' = [3 \quad 0 \quad 0 \quad 1 \quad 6 \quad 8]$.

Column vector: $\mathbf{y} = \begin{bmatrix} 6 \\ -9 \\ 13 \end{bmatrix}$.

Scalar: single number: 7 (1×1) .

Matrix of Matrices: $\mathbf{A} = \begin{bmatrix} \mathbf{P} & \mathbf{Q} & \mathbf{R} \\ \mathbf{X} & \mathbf{Y} & \mathbf{Z} \end{bmatrix}$, a partitioned matrix.

Complex numbers (seldom used in statistics):

$$\mathbf{M} = \begin{bmatrix} 2 + 3i & 7 + 9i \\ 8 - 6i & -5i \end{bmatrix} \qquad \bar{\mathbf{M}} = \begin{bmatrix} 2 - 3i & 7 - 9i \\ 8 + 6i & 5i \end{bmatrix} .$$

MATRIX ALGEBRA

- make up rules of the game that seem intuitively obvious, or motivated by matrices being arrays.

Addition:
$$\begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix} + \begin{bmatrix} 11 & 12 \\ 14 & -17 \end{bmatrix} = \begin{bmatrix} 1+11 & 3+12 \\ 2+14 & 6+(-17) \end{bmatrix} = \begin{bmatrix} 12 & 15 \\ 16 & -11 \end{bmatrix} .$$

Subtraction:
$$\begin{bmatrix} 13 \\ 21 \\ 16 \end{bmatrix} - \begin{bmatrix} 8 \\ 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 13-8 \\ 21-3 \\ 16-1 \end{bmatrix} = \begin{bmatrix} 5 \\ 18 \\ 15 \end{bmatrix} .$$

Zero(s):
$$\mathbf{0}_{2\times 2} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \qquad \mathbf{0}_{3\times 4} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} .$$

One(s):
$$\mathbf{I}_{2\times 2} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \qquad \mathbf{I}_{4\times 4} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

MULTIPLICATION

Method

Nutrition laboratory needs 20 rabbits and 150 mice.

Cost in Vancouver: \$5/rabbit, \$1/mouse.

Total cost: 20(5) + 150(1) = 250

$$[20 \quad 150]\begin{bmatrix}5\\1\end{bmatrix} = 20(5) + 150(1) = 250 \ .$$

Biochemistry needs 10 rabbits and 100 mice.

$$\begin{bmatrix}20 & 150\\10 & 100\end{bmatrix}\begin{bmatrix}5\\1\end{bmatrix} = \begin{bmatrix}20(5) + 150(1)\\10(5) + 100(1)\end{bmatrix} = \begin{bmatrix}250\\150\end{bmatrix} \ .$$

Costs in Seattle: \$3/rabbit \$½/mouse

$$\begin{bmatrix}20 & 150\\10 & 100\end{bmatrix}\begin{bmatrix}5 & 3\\1 & \frac{1}{2}\end{bmatrix} = \begin{bmatrix}20(5) + 150(1) & 20(3) + 150(\frac{1}{2})\\10(5) + 100(1) & 10(3) + 100(\frac{1}{2})\end{bmatrix} = \begin{bmatrix}250 & 135\\150 & 80\end{bmatrix} \ .$$

General Form

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times t} = \mathbf{P}_{r\times t} \ .$$

Multiplying by I

$$\mathbf{I}_{r\times r}\mathbf{A}_{r\times c} = \mathbf{A}_{r\times c} = \mathbf{A}_{r\times c}\mathbf{I}_{c\times c} \ .$$

Multiplying by a scalar

$$3\begin{bmatrix}1 & 6 & 2\\0 & 7 & 4\end{bmatrix} = \begin{bmatrix}3 & 18 & 6\\0 & 21 & 12\end{bmatrix} \ .$$

SOME PRODUCTS DON'T EXIST!

$6(4) = 24 = 4(6)$; always!

AX and **XA** can both be non-existent:

$$\mathbf{A} = [2 \quad 3] \qquad \mathbf{X} = [1 \quad 3 \quad 7] \; .$$

AX can exist, with **XA** non-existent (or vice versa):

$$\mathbf{A} = [2 \quad 3] \qquad \mathbf{X} = \begin{bmatrix} 1 & 3 & 7 \\ 2 & -1 & 4 \end{bmatrix} \; .$$

AX and **XA** can both exist, but not be equal:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 3 & 2 \\ 7 & 1 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 3 & 7 \\ 2 & -1 & 4 \end{bmatrix} \; .$$

AX and **XA** can both exist and be equal:

$$\mathbf{A} = \begin{bmatrix} 6 & 1 \\ 1 & 6 \end{bmatrix} \qquad \mathbf{X} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \; .$$

DIVISION: (Complicated!)

$2(\frac{1}{2}) = 1 = \frac{1}{2}(2)$ is $2(2^{-1}) = 1 = 2^{-1}(2)$

Analogously: $\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$

— but only for certain forms of **A**.

Application: linear equations

$$\begin{bmatrix} 3 & 8 & 1 \\ 1 & 1 & 1 \\ 2 & -1 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 22 \\ 6 \\ 15 \end{bmatrix} \quad \text{is} \quad \begin{aligned} 3x + 8y + z &= 22 \\ x + y + z &= 6 \\ 2x - y + 5z &= 15 \end{aligned}$$

$$\mathbf{At} = \mathbf{r}$$

$$\mathbf{A} = \begin{bmatrix} 3 & 8 & 1 \\ 1 & 1 & 1 \\ 2 & -1 & 5 \end{bmatrix} \qquad \mathbf{A}^{-1} = \frac{1}{9}\begin{bmatrix} -6 & 41 & -7 \\ 3 & -13 & 2 \\ 3 & -19 & 5 \end{bmatrix} \qquad \mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A}$$

$$\mathbf{A}^{-1}\mathbf{At} = \mathbf{A}^{-1}\mathbf{r}$$

$$\mathbf{It} = \mathbf{A}^{-1}\mathbf{r}$$

$$\mathbf{t} = \mathbf{A}^{-1}\mathbf{r} = \frac{1}{9}\begin{bmatrix} -6 & 41 & -7 \\ 3 & -13 & 2 \\ 3 & -19 & 5 \end{bmatrix}\begin{bmatrix} 22 \\ 6 \\ 15 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} .$$

CONTRADICTIONS WITH SCALAR ALGEBRA

Scalars: $ax = 0 \Rightarrow a = 0$ or $x = 0$

$ax = bx \Rightarrow (a - b)x = 0 \Rightarrow a = b$ or $x = 0$

Matrices: $\mathbf{AX} = \mathbf{0} \not\Rightarrow \mathbf{A} = \mathbf{0}$ or $\mathbf{X} = \mathbf{0}$

$$\begin{bmatrix} 6 & -4 \\ 3 & -2 \\ -18 & 12 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} .$$

$\mathbf{AX} = \mathbf{BX} \not\Rightarrow \mathbf{A} = \mathbf{B}$

$$\begin{bmatrix} 3 & 2 \\ 8 & 11 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 4 \\ 3 & 16 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} .$$

$\mathbf{A}^2 = \mathbf{A} \not\Rightarrow \mathbf{A} = \mathbf{I}$ or $\mathbf{0}$.

$$\begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} .$$

A FEW (OF MANY) SPECIAL OPERATIONS

Transposing

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 7 \\ 0 & 9 & 5 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 1 & 0 \\ 3 & 9 \\ 7 & 5 \end{bmatrix} .$$

Symmetry: $\mathbf{B} = \mathbf{B}'$

$$\mathbf{B} = \begin{bmatrix} 1 & 3 & 7 \\ 3 & 19 & -24 \\ 7 & -24 & 5 \end{bmatrix} \qquad \mathbf{B}' = \begin{bmatrix} 1 & 3 & 7 \\ 3 & 19 & -24 \\ 7 & -24 & 5 \end{bmatrix} .$$

Trace: sum of diagonal elements of a square matrix

$$\mathrm{tr}(\mathbf{B}) = 1 + 19 + 5 = 25 .$$

Square root: $\mathbf{K}$ is a value of $\sqrt{\mathbf{D}}$ if $\mathbf{K}^2 = \mathbf{D}$

$$\begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}^2 = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix}^2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} .$$

Determinant

$\det(\mathbf{A}_{n\times n})$: polynomial, order n, of the elements of $\mathbf{A}$.

Eigen roots (λ) and vectors ($\mathbf{w}$) $\qquad \mathbf{Aw} = \lambda\mathbf{w}$

$$\begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ -3 \end{bmatrix} = -5 \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 4 \\ 9 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = 7 \begin{bmatrix} 2 \\ 3 \end{bmatrix} .$$

A FEW (OF MANY) SPECIAL MATRICES

Orthogonal: $\mathbf{AA'} = \mathbf{A'A} = \mathbf{I}$.

Idempotent: $\mathbf{A}^2 = \mathbf{A}$

Scalar : $\lambda\mathbf{I}$

Triangular: $$\begin{bmatrix} 1 & 2 & 3 & 8 \\ 0 & 4 & 7 & 4 \\ 0 & 0 & 6 & 3 \\ 0 & 0 & 0 & 1 \end{bmatrix} .$$

Circulant : $$\begin{bmatrix} a & b & c & d \\ d & a & b & c \\ c & d & a & b \\ b & c & d & a \end{bmatrix} .$$

Band : $$\begin{bmatrix} a & b & \cdot & \cdot & \cdot \\ \cdot & a & b & \cdot & \cdot \\ \cdot & \cdot & a & b & \cdot \\ \cdot & \cdot & \cdot & a & b \\ \cdot & \cdot & \cdot & \cdot & a \end{bmatrix} .$$

Stochastic: $$\begin{bmatrix} .7 & .2 & .1 \\ .2 & .3 & .5 \\ .4 & .4 & .2 \end{bmatrix}$$

$0 \le$ every element ≤ 1
Elements of each row sum to 1.0 .

··· and on, and on.

3 Illustrations: economics, pest control, and genetics.

Illustration from economics

INPUT-OUTPUT ANALYSIS

Leontief, Wassily (1970) *Review of Economics and Statistics* **LII**, 262-271.

——— (1966) *Input-Output Economics.* Oxford.

Consider a society of 4 sectors:

	Agriculture	: total output	x_1
	Manufacturing	: total output	x_2
	Pollution elimination	: total effort	x_3
and	Labour	: labour force	x_4

To produce 1 unit of agricultural output (e.g., wheat production) suppose it requires

.25 units of agricultural output	– e.g., wheat seed
.14 units of manufacturing output	– e.g., combine-harvester
.50 units of pollution being produced	– e.g., diesel smoke
.80 units of labour	– e.g., combine driver

··· and so on

and y_1 = total consumer demand for agricultural output
y_2 = total consumder demand for manufacturing
y_3 = uneliminated pollution
y_4 = consumer demand for labour

Sector	Total Output	Requirements for producing one unit of				Final consumer demand
		Agric.	Mnfct.	Prv'n.	Labor	
Agriculture	x_1	.25	.40	0	0	y_1
Manufacturing	x_2	.14	.12	.2	0	y_2
Pollution prevention	x_3^*	.50	.20	0	0	$-y_3^*$
Labor	x_4	.80	3.60	2.0	0	y_4

* x_3 is total eliminated pollution.

* y_3 is total uneliminated pollution; hence $-y_3$ and x_3 are in same units.

We then have

$$\begin{aligned}
x_1 - (.25x_1 + .40x_2 \qquad\quad) &= y_1 \\
x_2 - (.14x_1 + .12x_2 + .2x_3) &= y_2 \\
x_3 - (.50x_1 + .20x_2 \qquad\quad) &= -y_3 \\
x_4 - (.80x_1 + 3.60x_2 + 2.0x_3) &= y_4
\end{aligned} \qquad \mathbf{x} - \mathbf{Ax} = \mathbf{y}$$

Sector	Total Output	Requirements for producing one unit of				Final consumer demand
		Agric.	Mnfct.	Prv'n.	Labor	
Agriculture	x_1	.25	.40	0	0	y_1
Manufacturing	x_2	.14	.12	.2	0	y_2
Pollution prevention	x_3^*	.50	.20	0	0	$-y_3^*$
Labor	x_4	.80	3.60	2.0	0	y_4

* x_3 is total eliminated pollution.

* y_3 is total uneliminated pollution; hence $-y_3$ and x_3 are in same units.

We then have

$$
\begin{aligned}
x_1 - (.25x_1 + .40x_2 \qquad\qquad) &= y_1 \\
x_2 - (.14x_1 + .12x_2 + .2x_3) &= y_2 \\
x_3 - (.50x_1 + .20x_2 \qquad\qquad) &= -y_3 \\
x_4 - (.80x_1 + 3.60x_2 + 2.0x_3) &= y_4
\end{aligned}
\qquad \mathbf{x} - \mathbf{A}\mathbf{x} = \mathbf{y}
$$

$$\begin{bmatrix} .75 & -.40 & 0 & 0 \\ -.14 & .88 & -.20 & 0 \\ .50 & .20 & -1.00 & 0 \\ -.80 & -3.60 & -2.00 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad (\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{y}$$

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1.573 & .749 & -.149 & 0 \\ .449 & 1.404 & -.281 & 0 \\ .876 & .655 & -1.131 & 0 \\ 4.628 & 6.965 & -3.393 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} \qquad \mathbf{x} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{y} \ .$$

Examples

[1] $$x_1 = 1.573y_1 + .749y_2 - .149y_3 \ ,$$

i.e., Agriculture output must equal

1.573(demand for agricultural products)
\+ .749(demand for manufactured goods)
\- .149(uneliminated pollution) .

[3] $$x_3 = .876y_1 + .655y_2 - 1.131y_3 \ ,$$

i.e., Pollution prevention effort must equal

.876(demand for agricultural products)
\+ .655(demand for manufactured goods)
\- 1.131(uneliminated pollution)

Note: $\uparrow$ uneliminated pollution $\downarrow$ outputs needed.

Illustration of pest control

RABBIT EXTERMINATION

Darwin, J.H. and Williams, R.M. (1964). The effect of time of hunting on the size of a rabbit population. *New Zealand Journal of Science* 7, 341-352.

In New Zealand and Australia

- rabbits were introduced from England, early 19*th* Century.
- in mild climates, fast became a pest.
- especially for sheep farmers.
- 6-8 rabbits eat as much grass as one sheep.
- extensive efforts at eradication.
- Australia: myxamatosis.
- New Zealand: poisoning.
- seasonal work: 1 man/20 farms.

Questions: When is best time of year to poison?

How far from optimum are other times?

Use of Leslie matrices

$$\mathbf{L} = \begin{bmatrix} \surd & \surd & \surd & \surd \\ \surd & \cdot & \cdot & \cdot \\ \cdot & \surd & \cdot & \cdot \\ \cdot & \cdot & \surd & \cdot \end{bmatrix} \leftarrow \text{reproductive rates, by age .}$$

survival rates, by age.

Measure age of rabbits in units of 4 weeks.

Rabbits mostly live 3 years or less.

Matrices $\mathbf{L}$ are 39×39.

One for each 4-week period of a year.

Distribution of rabbits by age 1-39:

$\mathbf{u}_0$ at start of year, 39 entries;

$\mathbf{u}_{13}$ at end of year;

$$\mathbf{u}_{13} = \mathbf{L}_1\mathbf{L}_2\cdots\mathbf{L}_{13}\mathbf{u}_0 \ .$$

Destruction policy: $\mathbf{D} = \begin{bmatrix} d_1 & \cdot & \cdot & & \\ \cdot & d_2 & \cdot & & \\ \cdot & \cdot & d_3 & & \\ & & & \ddots & \\ & & & & d_{39} \end{bmatrix}$: different killing rates for different ages.

Effect on $\mathbf{u}_{13}$

$$\left.\begin{array}{lll} \mathbf{M}_1 & = \mathbf{D}\mathbf{L}_1\mathbf{L}_2 & \cdots\ \mathbf{L}_{13} \\ \mathbf{M}_2 & = \mathbf{L}_1\mathbf{D}\mathbf{L}_2 & \cdots\ \mathbf{L}_{13} \\ \mathbf{M}_3 & = \mathbf{L}_1\mathbf{L}_2\mathbf{D} & \cdots\ \mathbf{L}_{13} \\ \vdots & & \vdots \\ \mathbf{M}_{14} & = \mathbf{L}_1\mathbf{L}_2 & \cdots\ \mathbf{L}_{13}\mathbf{D} \end{array}\right\} \mathbf{u}_{13}^{(i)} = \mathbf{M}_i\mathbf{u}_0 \rightarrow \lambda_i \text{ for } \mathbf{M}_i\mathbf{w} = \lambda_i\mathbf{w}\ .$$

λ_i is largest solution for λ to $\mathbf{M}_i\mathbf{w} = \lambda_i\mathbf{w}$
i.e., to the polynomial equation of order 39

$$\det(\mathbf{M}_i - \lambda\mathbf{I}) = 0\ .$$

λ_i is dominant eigenroot of $\mathbf{M}_i$

Under stable age distribution, λ_i represents proportional increase in rabbit population during one year, i.e., in $\mathbf{M}_i\mathbf{u}_0$ compared to $\mathbf{u}_0$.

And there were 11 different **D**-matrices: 14 **M**s and a λ for each.

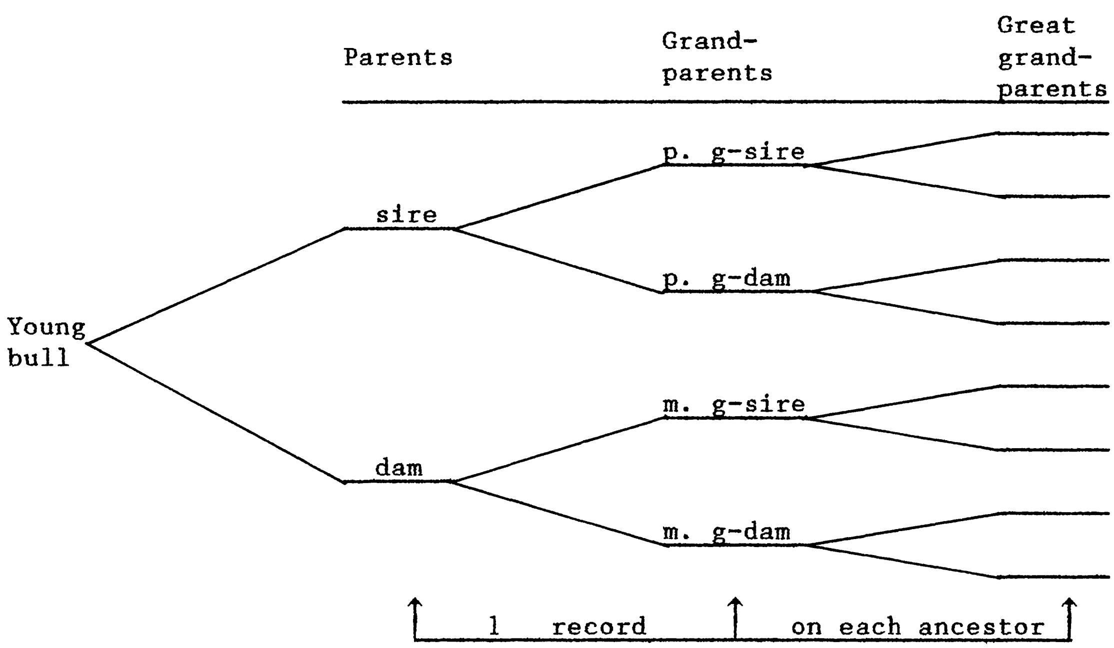
Result (not unexpected!) It is better to use a killing policy that kills more old than young rabbits just before the old start a new breeding season; and to use the opposite policy when the young have had a chance to be killed or die (naturally) but not much of a chance to breed, i.e., about six months away from the other optimum. Differences of 3 to 4 percent in the destruction rates showed appreciable differences in terms of overall effect in a population that is increasing but at not too fast a rate!

Illustration: genetics

CORRELATING ESTIMATED AND ACTUAL GENETIC VALUE

Searle, S.R. (1963). The efficiency of ancester records in animal breeding. *Heredity* **18**, 351-360.

<u>Question</u> How useful are records from parents, grandparents, great-grandparents ··· in estimating genetic worth of a new-born animal, e.g., a dairy bull?



Estimate genetic worth of young bull from 1 record on each ancestor.

g = true genetic worth of young bull

$\hat{g}_n$ = estimate of g using n generations of ancestors

$= b_1x_1 + b_2x_2 + \cdots + b_nx_n$

$= \sum_{i=1}^{n} b_ix_i$ for x_i mean record of i'th generation back

r_n = correlation of $\hat{g}_n$ and g

$$r_4^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \end{bmatrix} \begin{bmatrix} \frac{1}{2h} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{4} & \frac{1}{4h} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{8} & \frac{1}{8} & \frac{1}{8h} & \frac{1}{16} \\ \frac{1}{16} & \frac{1}{16} & \frac{1}{16} & \frac{1}{16}h \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{4} \\ \frac{1}{8} \\ \frac{1}{16} \end{bmatrix} .$$

Notation: h is a genetic parameter: $0 \le h \le 1$

Use k for $1/h$ and p for $\frac{1}{2}$

$$r_n^2 = [p \quad p^2 \quad p^3 \quad \cdots \quad p^n] \begin{bmatrix} kp & p^2 & p^3 & \cdots & p^n \\ p^2 & kp^2 & p^3 & \cdots & p^n \\ p^3 & p^3 & kp^3 & \cdots & p^n \\ \vdots & \vdots & \vdots & & \vdots \\ p^n & p^n & p^n & \cdots & kp^n \end{bmatrix}^{-1} \begin{bmatrix} p \\ p^2 \\ p^3 \\ \vdots \\ p^n \end{bmatrix}$$

$$= \frac{2h - 1}{2h} + \frac{(1 - h)^2}{2h(1 - hr_{n-1}^2)} .$$

Efficiency of single records of ancestors

Heritability h	Efficiency (r)		
	Upper Limit	Ancestors	
		Parents	Parents and grandparents
	$\sqrt{[h + \frac{1}{2} - \sqrt{(h+\frac{1}{2})^2-2h^2}]/2h}$	$r = \sqrt{\frac{1}{2}h}$	$r = \sqrt{\frac{1}{2}h(3-2h)/(2-h^2)}$
.1	.29	.22	.27
.3	.45	.39	.43
.5	.54	.50	.53
.7	.61	.59	.61
.9	.67	.67	.67
1.0	.71	.71	.71

Lecture II

SOME OF TODAY'S POPULAR MATRIX ALGEBRA

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

Theme: show some matrix operations,
that are not necessarily new
but nevertheless not found in many textbooks
and which are being used in applications.

1. Kronecker (direct) product
2. Generalized inverses
3. Inverting **A** + **UBV**
4. Vec and vech
5. Vec-permutation matrices
6. Alternatives to the spectral decomposition theorem.

Lecture III illustrates some uses of these in statistics.

1. KRONECKER (DIRECT) PRODUCTS

<u>Regular product</u>:

$$\begin{bmatrix} 1 & 7 \\ 2 & 5 \end{bmatrix}\begin{bmatrix} 6 & 3 & 0 \\ 4 & 2 & 9 \end{bmatrix} = \begin{bmatrix} 1(6) + 7(4) & 1(3) + 7(8) & 1(0) + 7(9) \\ 2(6) + 5(4) & 2(3) + 5(4) & 2(0) + 5(9) \end{bmatrix}$$

$$\mathbf{A}_{r\times c}\mathbf{B}_{c\times s} = \mathbf{P}_{r\times s} = \left\{_m \sum_{j=1}^{c} a_{ij}b_{jk}\right\}_{i=1\ k=1}^{r\quad s}$$

<u>Hadamard product</u>:

$$\begin{bmatrix} 1 & 7 \\ 2 & 5 \end{bmatrix} \cdot \begin{bmatrix} 6 & 3 \\ 4 & 8 \end{bmatrix} = \begin{bmatrix} 1(6) & 2(3) \\ 2(4) & 5(8) \end{bmatrix}$$

$$\mathbf{A}_{r\times c} \cdot \mathbf{B}_{r\times c} = \mathbf{P}_{r\times c} = \left\{_m\ a_{ij}b_{ij}\right\}_{i=1\ j=1}^{r\quad c}$$

<u>Kronecker product</u>:

$$\begin{bmatrix} 1 & 7 \\ 2 & 5 \end{bmatrix} * \begin{bmatrix} 6 & 3 & 0 \\ 4 & 8 & 9 \end{bmatrix} = \begin{bmatrix} 1\begin{bmatrix} 6 & 3 & 0 \\ 4 & 8 & 9 \end{bmatrix} & 7\begin{bmatrix} 6 & 3 & 0 \\ 4 & 6 & 9 \end{bmatrix} \\ & \\ 2\begin{bmatrix} 6 & 3 & 0 \\ 4 & 8 & 9 \end{bmatrix} & 5\begin{bmatrix} 6 & 3 & 0 \\ 4 & 8 & 9 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} 6 & 3 & 0 & 42 & 21 & 0 \\ 4 & 8 & 9 & 28 & 56 & 63 \\ 12 & 6 & 0 & 30 & 15 & 0 \\ 8 & 16 & 18 & 20 & 40 & 45 \end{bmatrix}$$

$$\mathbf{A}_{r\times c} * \mathbf{B}_{f\times g} = \mathbf{P}_{rf\times cg} = \left\{_m\ a_{ij}\mathbf{B}\right\}_{i=1\ j=1}^{r\quad c} .$$

Note: $\otimes$, not *, is the customary Kronecker product symbol.

PROPERTIES OF KRONECKER PRODUCT

Transpose : $(\mathbf{A} * \mathbf{B})' = \mathbf{A}' * \mathbf{B}'$

Contrast with $(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'$.

Vectors : $\mathbf{x}' * \mathbf{y} = \mathbf{y}\mathbf{x}' = \mathbf{y} * \mathbf{x}'$

Partitioning: $[\mathbf{A} \quad \mathbf{X}] * \mathbf{B} = [\mathbf{A} * \mathbf{B} \quad \mathbf{X} * \mathbf{B}]$

But $\mathbf{A} * [\mathbf{B} \quad \mathbf{Y}] \neq [\mathbf{A} * \mathbf{B} \quad \mathbf{A} * \mathbf{Y}]$

Products : $(\mathbf{A} * \mathbf{B})(\mathbf{X} * \mathbf{Y}) = (\mathbf{AX} * \mathbf{BY})$

— when all products exist.

<u>Inverse</u> : $(\mathbf{A} * \mathbf{B})^{-1} = \mathbf{A}^{-1} * \mathbf{B}^{-1}$: when $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$ exist.

Contrast with $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$

<u>Rank</u> : $r(\mathbf{A} * \mathbf{B}) = r(\mathbf{A})r(\mathbf{B})$

<u>Trace</u> : $tr(\mathbf{A} * \mathbf{B}) = tr(\mathbf{A})tr(\mathbf{B})$

<u>Determinant</u>: $\det(\mathbf{A} * \mathbf{B}) = 0$ for $\mathbf{A}_{r\times c}$ and $\mathbf{B}_{c\times r}$

$= (\det \mathbf{A})^q(\det \mathbf{B})^p$ for $\mathbf{A}_{p\times p}$ and $\mathbf{B}_{q\times q}$.

<u>Eigenroots</u> : root of $(\mathbf{A} * \mathbf{B})$ = (root of $\mathbf{A}$)(root of $\mathbf{B}$)

History

Review : Henderson, H.V., Pukelsheim, F. and Searle, S.R. (1983). *Linear and Multilinear Algebra* **14**, 113-120.

Origins: the result on determinants. Zehfuss, G. (1858). *Zeitschrift für Mathematik und Physik* **3**, 298-301.

Supported by Muir (1890-1923, 4 volumes), Rutherford (1933), Aitken (1935) and Ledermann (1936).

Kronecker ? His collected works were edited by Hensel, a student of Kronecker's, 1880-4, in Berlin: and Hensel (1889,1891) notes that Kronecker had given proof of the determinantal result. Subsequently *many* writers propounded the result as Kronecker's: and so his name came to be associated with the matrix operation $\mathbf{A} * \mathbf{B}$.

2. GENERALIZED INVERSE

Regular inverse

$$\mathbf{AA}^{-1} = \mathbf{I} = \mathbf{A}^{-1}\mathbf{A} \quad \text{only for } \begin{cases} \mathbf{A} \text{ square} \\ \det(\mathbf{A}) \neq 0 \ . \end{cases}$$

Moore-Penrose inverse

But for *any* $\mathbf{A}$ (other than $\mathbf{A} = \mathbf{0}$)

$$\left.\begin{array}{l} \mathbf{AMA} = \mathbf{A} \\ \mathbf{MAM} = \mathbf{M} \\ \mathbf{MA} \text{ symmetric} \\ \mathbf{AM} \text{ symmetric} \end{array}\right\} \mathbf{M} \text{ (Moore-Penrose inverse) is unique.}$$

For $\mathbf{A} = \mathbf{KL}$, with $\mathbf{K}$ ($\mathbf{L}$) of full column (row) rank, $r(\mathbf{A})$

$$\mathbf{M} = \mathbf{L}'(\mathbf{K}'\mathbf{AL}')^{-1}\mathbf{K}' \ .$$

Generalized inverse

Any $\mathbf{A}^-$ satisfying $\mathbf{AA}^-\mathbf{A} = \mathbf{A}^-$. $\quad \mathbf{A}_{p\times q}$ has $(\mathbf{A}^-)_{q\times p}$.

Many generalized inverses

Given one $\mathbf{A}^-$, another is

$$\mathbf{G}_1 = \mathbf{A}^-\mathbf{AA}^- + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{T} + \mathbf{S}(\mathbf{I} - \mathbf{AA}^-) \text{ for any } \mathbf{S} \text{ and } \mathbf{T} ;$$

and

$$\mathbf{G}_2 = \mathbf{A}^- + \mathbf{U} - \mathbf{A}^-\mathbf{AUAA}^- \text{ for any } \mathbf{U} .$$

Proof: Suppose $\mathbf{AG}^*\mathbf{A}$ for $\mathbf{G}^*$ different from $\mathbf{A}^-$.

Then $\mathbf{T} = \mathbf{G}^*$ and $\mathbf{S} = \mathbf{A}^-\mathbf{AG}^* \quad \Rightarrow \quad \mathbf{G}_1 = \mathbf{G}^*$

and $\mathbf{U} = \mathbf{G}^* - \mathbf{A}^- + \mathbf{A}^-\mathbf{AA}^- \quad \Rightarrow \quad \mathbf{G}_2 = \mathbf{G}^*$.

Deriving **M**

$$\mathbf{M} = \mathbf{A}'(\mathbf{AA}')^-\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'$$

Note: $\mathbf{A}(\mathbf{A}'\mathbf{A})^-\mathbf{A}'\mathbf{A} = \mathbf{A}$ for real $\mathbf{A}$.

[Use $\mathbf{PA}'\mathbf{A} = \mathbf{QA}'\mathbf{A} \Rightarrow \mathbf{PA}' = \mathbf{QA}'$]

Special cases

Non-singular $\mathbf{A}$:	$\mathbf{A}^- = \mathbf{A}^{-1}$, only.	Orthogonal $\mathbf{A}$:	$\mathbf{A}^- = \mathbf{A}'$
Idempotent $\mathbf{A}$:	$\mathbf{A}^- = \mathbf{A}$	Null $\mathbf{A}$:	$\mathbf{A}^-$ = anything.

2a. SOLVING LINEAR EQUATIONS

Consistent equations (solution exists): $\mathbf{Ax} = \mathbf{y}$ for $\mathbf{y} \neq \mathbf{0}$.

Solution: $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ when $\mathbf{A}^{-1}$ exists.

$$\mathbf{x} = \mathbf{A}^{-}\mathbf{y} \begin{cases} \text{when } \mathbf{A} \text{ is square, without } \mathbf{A}^{-1} \text{ existing} \\ \text{when } \mathbf{A} \text{ is rectangular} \end{cases}$$

if and only if $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$. (Rao, 1963)

Proof (i) $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A} \Rightarrow \mathbf{AA}^{-}\mathbf{Ax} = \mathbf{Ax}$

$\mathbf{Ax} = \mathbf{y} \Rightarrow \mathbf{A}(\mathbf{A}^{-}\mathbf{y}) = \mathbf{y} \Rightarrow \mathbf{A}^{-}\mathbf{y}$ is a solution

(ii) Equations $\mathbf{Ax} = \mathbf{a}_j$, the j'th column of $\mathbf{A}$, have a solution $\mathbf{x} = \mathbf{e}_j$, the j'th column of $\mathbf{I}$: e.g.,

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a_{12} \\ a_{22} \end{bmatrix} .$$

Therefore equations $\mathbf{Ax} = \mathbf{a}_j$ are consistent .

But so are $\mathbf{Ax} = \mathbf{y}$, with solution $\mathbf{x} = \mathbf{A}'\mathbf{y}$.

Therefore $\mathbf{Ax} = \mathbf{a}_j$ have solution $\mathbf{x} = \mathbf{A}^{-}\mathbf{a}_j$.

Hence $\mathbf{AA}^{-}\mathbf{a}_j = \mathbf{a}_j$; i.e., $\mathbf{AA}^{-}\mathbf{A} = \mathbf{A}$.

A general solution

Theorem $\mathbf{x} = \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z}$ for arbitrary $\mathbf{z}$.

generates *all* solutions.

Proof: Suppose x^* is a solution.

Then $\mathbf{z} = -\mathbf{x}^* \Rightarrow \mathbf{x} = \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{x}^* = \mathbf{A}^{-}\mathbf{y} + \mathbf{x}^* - \mathbf{A}^{-}\mathbf{y} - \mathbf{x}^*$.

Theorem $\mathbf{x} = \mathbf{A}^{-}\mathbf{y}$ for *all* $\mathbf{A}^{-}$ generates all solutions.

Lemma For arbitrary $\mathbf{z}$ and known $\mathbf{y}$ there exists arbitrary $\mathbf{X}$ such that $\mathbf{z} = \mathbf{Xy}$,

$$\text{e.g., } \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & z_1/y_2 & 0 \\ 0 & z_2/y_2 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} .$$

Proof

$$\begin{aligned} \mathbf{x} &= \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z} \\ &= \mathbf{A}^{-}\mathbf{y} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{Xy} \\ &= [\mathbf{A}^{-}\mathbf{A}\mathbf{A}^{-} + (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{X} + \mathbf{A}^{-}(\mathbf{I} - \mathbf{A}\mathbf{A}^{-})]\mathbf{y} \end{aligned}$$

$= \mathbf{G}_1\mathbf{y}$; and $\mathbf{G}_1$ can be *all* the $\mathbf{A}^{-}$ matrices.

Combinations of solutions to $\mathbf{Ax} = \mathbf{y} \neq \mathbf{0}$

For solutions $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_t$

$$\sum_{i=1}^{t} \lambda_i \mathbf{x}_i \quad \text{is a solution iff} \quad \Sigma\lambda_i = 0 \ .$$

Linearly independent solutions

$\mathbf{A}_{n \times q}$, rank $r_{\mathbf{A}}$ => $q - r_{\mathbf{A}} + 1$ LIN solutions .

An invariance property

For $\tilde{\mathbf{x}}$ being a solution to $\mathbf{Ax} = \mathbf{y} \neq 0$

$\mathbf{k}'\tilde{\mathbf{x}}$ is invariant to $\tilde{\mathbf{x}}$ iff $\mathbf{k}' = \mathbf{k}'\mathbf{A}^{-}\mathbf{A}$.

Equations $\mathbf{Ax} = \mathbf{0}$

Solutions $\mathbf{x} = (\mathbf{I} - \mathbf{A}^{-}\mathbf{A})\mathbf{z}$, arbitrary $\mathbf{z}$.

Solutions $\tilde{\mathbf{x}}_1, \cdots, \tilde{\mathbf{x}}_t$ => $\Sigma\lambda_i\tilde{\mathbf{x}}_i$ is a solution .

$\mathbf{A}_{n \times q}$, rank $r_{\mathbf{A}}$ => $q - r_{\mathbf{A}}$ LIN solutions .

Invariance: $\mathbf{k}'\tilde{\mathbf{x}} = 0$ for all $\tilde{\mathbf{x}}$ and all $\mathbf{k}' = \mathbf{k}'\mathbf{A}^{-}\mathbf{A}$.

3. $(A + UBV)^{-1}$

A long history: reviewed in Henderson and Searle (1981), *Siam Review* **23**, 53-60

(i) Determinant of a bordered matrix. Cauchy, 1789-1859, and Darboux (1874).

Frobenius (1908): $\begin{vmatrix} \mathbf{A} & \mathbf{u} \\ \mathbf{v}' & d \end{vmatrix} = d|\mathbf{A}| - \mathbf{v}'(\text{adj } \mathbf{A})\mathbf{u}$.

Schur (1917): $\begin{vmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{vmatrix} = |\mathbf{A}|\,|\mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{U}|$.

He used
$$\begin{bmatrix} \mathbf{A}^{-1} & \mathbf{0} \\ \mathbf{V}\mathbf{A}^{-1} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{A}^{-1}\mathbf{U} \\ \mathbf{0} & \mathbf{D} - \mathbf{V}\mathbf{A}^{-1}\mathbf{U} \end{bmatrix} \qquad (*)$$

(ii) $\begin{bmatrix} \mathbf{A} & \mathbf{U} \\ \mathbf{V} & \mathbf{D} \end{bmatrix}^{-1}$: is readily available from (*) .

Shur (1917), Boltz (1923) and Aitken (1934) were close.

Banachiewicz (1937) and Stankiewicz (1938).

Hotelling (1943), Duncan (1944) and Aitken (1946):
various other forms requiring $\mathbf{A}^{-1}$ and $\mathbf{D}^{-1}$.

(iii) $(\mathbf{A} + \mathbf{UBV})^{-1}$

Sherman and Morrison (1949,1950): altered a row or column.

Bartlett (1951): $\mathbf{A} + \mathbf{uv}' = \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{uv}'\mathbf{A}^{-1}/(1 + \mathbf{v}'\mathbf{A}^{-1}\mathbf{u})$.

Woodbury (1950): generalized to $(\mathbf{A} + \mathbf{UBV})^{-1}$ – but required $\mathbf{A}^{-1}$ and $\mathbf{B}^{-1}$.

Householder (1953): sometimes (erroneously) attributed with results.

Ouellette (1978): lengthy discussion.

(iv) $(\mathbf{A} + \mathbf{UBU}')^{-1}$: symmetry.

Guttman (1940): with $\mathbf{B} = \mathbf{I}$.

Henderson *et al.* (1959): in a statistical context.

Lindley and Smith (1972): "unexpected by-product" of a probabilistic argument.

Harville (1977): a form not requiring $\mathbf{B}^{-1}$.

<u>Broader results</u> (Henderson & Searle, 1981)

$(\mathbf{A} + \mathbf{UBV})$ with $\mathbf{A}^{-1}$ existing and $\mathbf{B}$ not necessarily square .

Utilize, with $(\mathbf{I} + \mathbf{P})^{-1}$ existing, that $\mathbf{I} = \mathbf{I} + \mathbf{P} - \mathbf{P}$ implies

$$(\mathbf{I} + \mathbf{P})^{-1} = \mathbf{I} - \mathbf{P}(\mathbf{I} + \mathbf{P})^{-1} = \mathbf{I} - (\mathbf{I} + \mathbf{P})^{-1}\mathbf{P} . \qquad [1]$$

Also, $$\mathbf{P}(\mathbf{I} + \mathbf{QP}) = (\mathbf{I} + \mathbf{PQ})\mathbf{P}$$

implies

$$(\mathbf{I} + \mathbf{PQ})^{-1}\mathbf{P} = \mathbf{P}(\mathbf{I} + \mathbf{QP})^{-1} . \qquad [2]$$

Then $$\begin{aligned}(\mathbf{A} + \mathbf{UBV})^{-1} &= (\mathbf{I} + \mathbf{A}^{-1}\mathbf{UBV})^{-1}\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - (\mathbf{I} + \mathbf{A}^{-1}\mathbf{UBV})^{-1}\mathbf{A}^{-1}\mathbf{UBVA}^{-1} \quad \text{from (1]} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}(\mathbf{I} + \mathbf{UBVA}^{-1})^{-1}\mathbf{UBVA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{U}(\mathbf{I} + \mathbf{BVA}^{-1}\mathbf{U})^{-1}\mathbf{BVA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UB}(\mathbf{I} + \mathbf{VA}^{-1}\mathbf{UB})^{-1}\mathbf{VA}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBV}(\mathbf{I} + \mathbf{A}^{-1}\mathbf{UBV})^{-1}\mathbf{A}^{-1} \\ &= \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{UBVA}^{-1}(\mathbf{I} + \mathbf{UBVA}^{-1})^{-1} ,\end{aligned}$$

all these last 5 expressions coming from [2].

Generalized inverses: $(\mathbf{A} + \mathbf{UBV})^-$

Harville (1977): $\mathbf{A}$ and $\mathbf{UBV}$ symmetric

$C(\mathbf{UBV}) \subset C(\mathbf{A})$

Henderson and Searle (1981):

$C(\mathbf{UBV}) \subset C(\mathbf{A})$, i.e., $\mathbf{AA^-UBV} = \mathbf{UBV}$

$R(\mathbf{UBV}) \subset R(\mathbf{A})$, i.e., $\mathbf{UBVA^-A} = \mathbf{UBV}$

Five forms of $(\mathbf{A} + \mathbf{UBV})^-$.

$$\mathbf{G}_1 = \mathbf{A}^- - \mathbf{A}^-(\mathbf{A}^- + \mathbf{A^-UBVA^-})^-\mathbf{A^-UBVA^-}$$

$$\mathbf{G}_2 = \mathbf{A}^- - \mathbf{A^-U}(\mathbf{U} + \mathbf{UBVA^-U})^-\mathbf{UBVA^-}$$

$$\mathbf{G}_3 = \mathbf{A}^- - \mathbf{A^-UB}(\mathbf{B} + \mathbf{BVA^-UB})^-\mathbf{BVA^-}$$

$$\mathbf{G}_4 = \mathbf{A}^- - \mathbf{A^-UBV}(\mathbf{V} + \mathbf{VA^-UBV})^-\mathbf{VA^-}$$

$$\mathbf{G}_5 = \mathbf{A}^- - \mathbf{A^-UBVA^-}(\mathbf{A}^- + \mathbf{A^-UBVA^-})^-\mathbf{A}^-$$

Note: When $\mathbf{A} + \mathbf{UBV}$ is non-singular, but $\mathbf{A}$ is singular, then

$$(\mathbf{A} + \mathbf{UBV})^{-1} = (\mathbf{A} + \mathbf{UBV})^- = \mathbf{G}_i \qquad \text{for } i = 1,\cdots,5 .$$

4. VEC AND VECH OPERATORS

vec(**A**): stacks columns of **A** in a single column.

$$\operatorname{vec}\begin{bmatrix} 1 & 17 & -4 \\ 2 & 29 & 5 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 17 \\ 29 \\ -4 \\ 5 \end{bmatrix} .$$

<u>Origins</u>: Sylvester (1884), Roth (1924), Koopmans (1950).

<u>Many names</u>: stack, pack, column-rolled out, column string

<u>Variations</u>: vec(**A**'): rows of **A**, as columns, stacked

[vec(**A**')]': rows of **A**, alongside one another

Vech: for symmetric matrices

$$\text{vec}\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} a \\ b \\ b \\ c \end{bmatrix} \qquad \text{vech}\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix} .$$

Extensions: to other patterned matrices

$$\text{veck}\begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$$

vecp: for any patterned matrix;
e.g., circulant, centrosymmetric, triangular, ...

vech**A** is a subset of vec**A**

$$\text{vech}\mathbf{A} = \mathbf{H}\text{vec}\mathbf{A} \qquad \text{vec}\mathbf{A} = \mathbf{G}\text{vech}\mathbf{A}$$

$$\Rightarrow \text{vech}\mathbf{A} = \mathbf{H}\mathbf{G}\text{vech}\mathbf{A} \quad \Rightarrow \quad \mathbf{H}\mathbf{G} = \mathbf{I} \quad \Rightarrow \quad \mathbf{H} = (\mathbf{G}'\mathbf{G})^{-1}\mathbf{G}' .$$

$$|\mathbf{H}(\mathbf{A} * \mathbf{A})\mathbf{G}| = |\mathbf{A}|^{n+1} \qquad \text{for } \mathbf{A}_{n\times n}$$

<u>vec of a product</u>

$$\mathrm{vec}(\mathbf{xy}') = \mathbf{y} * \mathbf{x};\ \text{by definition of vec and of } * .$$

$$\mathrm{vec}(\mathbf{Axy}'\mathbf{C}) = \mathbf{C}'\mathbf{y} * \mathbf{Ax} = (\mathbf{C}' * \mathbf{A})(\mathbf{y} * \mathbf{x}) = (\mathbf{C}' * \mathbf{A})\mathrm{vec}(\mathbf{xy}') .$$

Define $\mathbf{e}_i \equiv$ i'th column of $\mathbf{I}$ so that

$$\mathbf{B} = \{b_{ij}\} = \Sigma_i\Sigma_j b_{ij}\mathbf{e}_i\mathbf{e}_j' .$$

Then

$$\begin{aligned}\mathrm{vec}(\mathbf{ABC}) &= \mathrm{vec}(\mathbf{A}\ \Sigma_i\Sigma_j b_{ij}\mathbf{e}_i\mathbf{e}_j'\mathbf{C}) = \Sigma_i\Sigma_j b_{ij}\mathrm{vec}(\mathbf{Ae}_i\mathbf{e}_j'\mathbf{C}) \\ &= \Sigma_i\Sigma_j b_{ij}(\mathbf{C}' * \mathbf{A})\mathrm{vec}(\mathbf{e}_i\mathbf{e}_j') = (\mathbf{C}' * \mathbf{A})\mathrm{vec}(\Sigma_i\Sigma_j b_{ij}\mathbf{e}_i\mathbf{e}_j') \\ &= (\mathbf{C}' * \mathbf{A})\mathrm{vec}\mathbf{B} .\end{aligned}$$

$$\mathrm{vec}(\mathbf{AB}) = (\mathbf{I} * \mathbf{A})\mathrm{vec}\mathbf{B} = (\mathbf{B}' * \mathbf{A})\mathrm{vec}\mathbf{I} = (\mathbf{B}' * \mathbf{I})\mathrm{vec}\mathbf{A}$$

$$\begin{aligned}\mathbf{B} = \mathbf{A}^{-1} \quad \Rightarrow \quad \mathrm{vec}\mathbf{A}^{-1} &= (\mathbf{I} * \mathbf{A})^{-1}(\mathbf{A}^{-1'} * \mathbf{I})\mathrm{vec}\mathbf{A} \\ &= (\mathbf{A}^{-1'} * \mathbf{A}^{-1})\mathrm{vec}\mathbf{A} .\end{aligned}$$

Traces: $\mathrm{tr}(\mathbf{A}'\mathbf{B}) = (\mathrm{vec}\mathbf{A})'\mathrm{vec}\mathbf{B}$.

$$\mathrm{tr}(\mathbf{AX}'\mathbf{BXC}) = (\mathrm{vec}\mathbf{X})'(\mathbf{A}'\mathbf{C}' * \mathbf{B})\mathrm{vec}\mathbf{X} = (\mathrm{vec}\mathbf{X})'(\mathbf{CA} * \mathbf{B}')\mathrm{vec}\mathbf{X} .$$

Equations: $\mathbf{AXB} = \mathbf{C}$

1. $(\mathbf{B}' * \mathbf{A})\mathrm{vec}\mathbf{X} = \mathrm{vec}\mathbf{C}$

$$\mathrm{vec}\mathbf{X} = (\mathbf{B}' * \mathbf{A})^{-}\mathrm{vec}\mathbf{C} = (\mathbf{B}'^{-} * \mathbf{A}^{-})\mathrm{vec}\mathbf{C} .$$

2. $\Sigma_i \mathbf{A}_i\mathbf{X}\mathbf{B}_i + \Sigma_j \mathbf{C}_j\mathbf{X}'\mathbf{D}_j = \mathbf{K}$

$$\left\{\Sigma_i(\mathbf{B}_i' * \mathbf{A}) + [\Sigma_j(\mathbf{D}_j' * \mathbf{C}_j)]\mathbf{I}_{n,m}\right\}\mathrm{vec}\mathbf{X} = \mathrm{vec}\mathbf{K} .$$

– for $\mathbf{X}_{n\times m}$ and for $\mathbf{I}_{n,m}$ a vec-permutation matrix .

Jacobians

$$\mathbf{x} \to \mathbf{y} \qquad \mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}'} = \left\{\frac{\partial x_i}{\partial y_j}\right\} .$$

$$\mathbf{X} \to \mathbf{Y} \qquad \mathbf{J} = \frac{\partial \mathrm{vec}\mathbf{X}}{\partial(\mathrm{vec}\mathbf{Y})'}$$

e.g.: $\mathbf{Y} = \mathbf{AXB} \Rightarrow \mathrm{vec}(\mathbf{Y}) = (\mathbf{B}' * \mathbf{A})\mathrm{vec}\mathbf{X}$.

$$|\mathbf{J}| = 1/|\mathbf{B}' * \mathbf{A}| = |\mathbf{B}|^{-a}|\mathbf{A}|^{-b}, \text{ for } \mathbf{A}_{a\times a} \text{ and } \mathbf{B}_{b\times b} .$$

Henderson and Searle, 1979, *Canadian Journal of Statistics* **7**, 65-81.

5. VEC-PERMUTATION MATRICES

$\left.\begin{array}{l}\mathrm{vec}\mathbf{A}\\ \mathrm{vec}\mathbf{A}'\end{array}\right\}$ obviously have the same elements – in different orders

$\Rightarrow$ $\mathrm{vec}\mathbf{A} = \mathbf{P}\mathrm{vec}\mathbf{A}'$: $\mathbf{P}$ is a permutation matrix (I with rows permuted).

$$\mathbf{A} = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \qquad \mathbf{A}' = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

$$\mathrm{vec}\mathbf{A} \begin{bmatrix} 1 \\ 4 \\ 2 \\ 5 \\ 3 \\ 6 \end{bmatrix} = \begin{bmatrix} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{bmatrix} = \mathbf{P}\mathrm{vec}\mathbf{A}'$$

$\mathrm{P} = \left\{ \begin{array}{l} \mathrm{I}_{pq \times pq} \text{ with rows in order: } 1,\ 1+q,\ \cdots\ 1+(p-1)q, \\ \qquad 2,\ 2+q,\ \cdots\ 2+(p-1)q,\ 3,\ \cdots\ q,\ 2q,\ \cdots,\ pq\ . \end{array} \right.$

$$\mathbf{P} = \left[\begin{array}{ccc|ccc} 1 & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \hline \cdot & 1 & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 1 & \cdot \\ \hline \cdot & \cdot & 1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & 1 \end{array}\right] = \mathbf{I}_{2,3}$$

$$\mathrm{vec}\mathbf{A}_{p\times q} = \mathbf{I}_{p,q}\ \mathrm{vec}(\mathbf{A}')_{q\times p}$$

$\mathbf{I}_{p,q}$: $\left\{ \begin{array}{l} \text{order } pq \times pq \\ q \text{ rows and } p \text{ columns of sub-matrices} \\ \text{each sub-matrix is null except its } (i,j)\text{'th element is unity .} \end{array} \right.$

Properties

$$\mathbf{I}_{p,1} = \mathbf{I}_{1,p} = \mathbf{I}_{p\times p}; \quad (\mathbf{I}_{p,q})' = \mathbf{I}_{q,p}; \quad \mathbf{I}_{p,q}\mathbf{I}_{q,p} = \mathbf{I}_{pq\times pq}$$

$$\mathrm{vec}\mathbf{A}_{p\times q} = \mathbf{I}_{p,q}\,\mathrm{vec}(\mathbf{A}')_{q\times p} \quad \text{and} \quad \mathrm{vec}(\mathbf{A}')_{q\times p} = \mathbf{I}_{q,p}\,\mathrm{vec}\mathbf{A}_{p\times q}$$

$$\mathbf{B}_{p\times q} * \mathbf{A}_{r\times s} = \mathbf{I}_{r,p}(\mathbf{A}_{r\times s} * \mathbf{B}_{p\times q})\mathbf{I}_{q,s}$$

$$\det(\mathbf{I}_{p,q}) = (-1)^{p(p-1)q(q-1)/4}$$

$$\mathrm{tr}(\mathbf{I}_{p,q}) = 1 + \text{largest common factor of } p - 1 \text{ and } q - 1 \;.$$

Origins: numerous, varied, and with many names.

Henderson and Searle (1981), *Linear and Multilinear Algebra* **9**, 271-288.

Use and extensions: *loc cit* and

Neudecker and Wansbeek (1983), *Canadian Journal of Statistics* **11**, 221-231.

6. ALTERNATIVE TO THE SPECTRAL DECOMPOSITION THEOREM

A square and real:

$$\mathbf{AU} = \mathbf{UD}, \text{ for } \mathbf{D} = \mathrm{diag}\{\lambda_i, \text{eigenroots}\}; \quad \forall\, \mathbf{A} .$$

A simple: $\mathbf{U}^{-1}$ exists .

Spectral decomposition (SD) theorem

For **A** simple and non-singular with $\lambda_1, \cdots, \lambda_s$ distinct, there exist $\mathbf{M}_1, \cdots, \mathbf{M}_s$ such that

$$\mathbf{A} = \Sigma\lambda_t M_t \quad \text{and} \quad \mathbf{A}^{-1} = \Sigma\lambda_t^{-1}\mathbf{M}_t$$

where

$$\Sigma\mathbf{M}_t = \mathbf{I}, \quad \mathbf{M}_t = \mathbf{M}_t^2 \quad \text{and} \quad \mathbf{M}_t\mathbf{M}_{t'} = \mathbf{0} \quad \text{for } t \neq t' .$$

<u>Linear combinations of matrices</u> (statistics)

$$\mathbf{V} = \sum_{i=1}^{c} \theta_i \mathbf{K}_i \ .$$

$\mathbf{K}_i$: linearly independent and simultaneously diagonable

$$\mathbf{D}^{-1}\mathbf{K}_i\mathbf{P} = \mathbf{D}_i \qquad \forall\ \mathbf{K}_i$$

$$\mathbf{e}(\mathbf{K}_i) = \text{vector of eigenroots of } \mathbf{K}_i = \mathbf{e}(\mathbf{D}_i) = \mathbf{D}_i\mathbf{1} \ .$$

Often need $\mathbf{V}^{-1}$; => need $\mathbf{e}(\mathbf{V})$ for SD theorem.

Frequently $\mathbf{e}(\mathbf{K}_i)$ is easier to obtain than $\mathbf{e}(\mathbf{V})$.

Two theorems connecting $\mathbf{e}(\mathbf{V})$ to $\mathbf{e}(\mathbf{K}_i)$.

Theorem 1 (Eigenroots)

Define $\mathbf{L}_c = \left\{_r\ \mathbf{e}(\mathbf{K}_i)\right\}_{i=1}^{c}$; of full column rank.

Then $\mathbf{e}(\mathbf{V}) = \mathbf{L}_c\boldsymbol{\theta}_c$.

Theorem 2 (Inverse)

Suppose $\{\mathbf{K}_i : i = 1, \cdots, c\} \subset \{\mathbf{K}_i : i = 1, \cdots, q \geq c\}$
$\downarrow$
closed under multiplication

$$\mathbf{V} = \sum_{i=1}^{q} \theta_i \mathbf{K}_i \; ; \quad \text{some } \theta_i\text{-values zero}$$

$$\mathbf{e}(\mathbf{V}) = \mathbf{L}_q \boldsymbol{\theta}_q \; ; \quad \text{from Theorem 1}$$

$$= [\delta_1 \cdots \delta_n]', \text{ say .}$$

$$\mathbf{e}(\mathbf{V}^{-1}) = [1/\delta_1 \cdots 1/\delta_n]' \; .$$

Then $$\mathbf{V}^{-1} = \sum_{i=1}^{q} \tau_i \mathbf{K}_i \quad \text{with} \quad \boldsymbol{\tau}_q = (\mathbf{L}_q'\mathbf{L}_q)^{-1}\mathbf{L}_q'[\mathbf{e}(\mathbf{V}^{-1})] \; .$$

Henderson and Searle (1989). *Linear Algebra and Applications*
— Special issue on statistics.

Applications:

$\mathbf{K}_i$: commutative.

$\mathbf{K}_i$: $\mathbf{A}^i$, powers of a matrix; e.g., $\mathbf{V}$ a circulant.

$\mathbf{K}_i$: Kronecker products of powers;
e.g., certain covariance matrices.

Lecture III

USING MATRIX ALGEBRA IN STATISTICS

Shayle R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

STATISTICS

$\bar{x}$ and s^2

Data: $x_1 \quad x_2 \cdots x_n : \mathbf{x} = [x_1 \quad x_2 \cdots x_n]'$.

Make inferences to a population from which $\mathbf{x}$ represents a random sample.

$$\bar{x} = \sum_{i=1}^{n} x_i/n = \mathbf{x}'\mathbf{1}_n$$

$$s^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \mathbf{x}'(\mathbf{I} - \bar{\mathbf{J}}_n)\mathbf{x} .$$

1 and J

$\mathbf{1}'$ is a summing vector: e.g., $\mathbf{1}_3' = [1 \quad 1 \quad 1]$, and $\mathbf{1}_3'\mathbf{x} = x_1 + x_2 + x_3$

$\mathbf{J}_n = \mathbf{1}_n\mathbf{1}_n' \quad (\mathbf{1}_n'\mathbf{1}_n = n)$; and $\bar{\mathbf{J}}_n = \frac{1}{n}\mathbf{J}_n$

$$\text{e.g., } \mathbf{J}_3 = \mathbf{1}_3\mathbf{1}_3' = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} ; \quad \bar{\mathbf{J}}_3 = \frac{1}{3}\mathbf{J}_3 .$$

$$\mathbf{J}_n^2 = n\mathbf{J}_n \quad \text{and} \quad \bar{\mathbf{J}}_n^2 = \bar{\mathbf{J}}_n .$$

MEAN AND VARIANCE OF RANDOM VARIABLES

Consider $x_1, x_2, \cdots, x_n$ as random variables:

$$E(x_i) = \mu_i \quad \Rightarrow \quad E(\mathbf{x}) = \boldsymbol{\mu} = \{\mu_i\}$$

$$= \mu\mathbf{1} \quad \text{for} \quad \mu_i = \mu \ \forall \ i$$

$$v(x_i) = E(x_i - \mu_i)^2$$

$$\text{cov}(x_i, x_j) = E(x_i - \mu_i)(x_j - \mu_j) = E(x_j - \mu_j)(x_i - \mu_i)$$

$$\mathbf{V} = \text{var}(\mathbf{x}) = \begin{bmatrix} v(x_1) & \text{cov}(x_1,x_2) & \cdots & \text{cov}(x_1,x_n) \\ \text{cov}(x_2,x_1) & v(x_2) & \cdots & \text{cov}(x_2,x_n) \\ \vdots & & \ddots & \\ \text{cov}(x_n,x_1) & \cdots & & v(x_n) \end{bmatrix}$$

$$= E[(\mathbf{x} - \boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})']$$

$$= E(\mathbf{x}\mathbf{x}') - \boldsymbol{\mu}\boldsymbol{\mu}' \ .$$

$$\text{c.f.,} \ v(x_i) = E(x_i^2) - \mu_i^2 \ .$$

NORMALITY

One variable

$$x \sim N(\mu,\sigma^2)$$

$$f(x) \quad = \frac{e^{-\frac{1}{2}(x - \mu_i)^2/\sigma^2}}{(2\pi\sigma^2)^{\frac{1}{2}}}$$

$$\text{m.g.f.} \equiv M_{x_i}^{(t)} = E(e^{xt}) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} .$$

$$\mu_x^{(k)} \quad = E(x^k) = \left.\frac{\partial M_x^{(t)}}{\partial t}\right|_{t=0} .$$

n variables

$$\mathbf{x} \sim N(\boldsymbol{\mu},\mathbf{V})$$

$$f(\mathbf{x}) = \frac{e^{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{x} - \boldsymbol{\mu})}}{(2\pi)^{n/2}|\mathbf{V}|^{\frac{1}{2}}}$$

$$M_{\mathbf{x}}^{(t)} = E(e^{\mathbf{t}'\mathbf{x}}) = e^{\boldsymbol{\mu}'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{V}\mathbf{t}}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \qquad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} \qquad \mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} .$$

Marginal: $\mathbf{x}_1 \sim N(\boldsymbol{\mu}_1,\mathbf{V}_{11})$

Conditional: $\mathbf{x}_1|\mathbf{x}_2 \sim N\left[\boldsymbol{\mu}_1 + \mathbf{V}_{12}\mathbf{V}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2),\quad \mathbf{V}_{11} - \mathbf{V}_{12}\mathbf{V}_{22}^{-1}\mathbf{V}_{12}\right]$.

LINEAR TRANSFORMATIONS

$$E(\mathbf{Tx}) = \mathbf{T}\boldsymbol{\mu} \quad \text{and} \quad \mathrm{var}(\mathbf{Tx}) = \mathbf{TVT}' \ .$$

QUADRATIC FORMS

Basics

$\mathbf{x}'\mathbf{A}\mathbf{x}$ with $\mathbf{A} = \mathbf{A}'$, and $\mathbf{x} \sim N(\boldsymbol{\mu}, \mathbf{V})$, with $\mathbf{V}^{-1}$ existing .

$$E(\mathbf{x}'\mathbf{A}\mathbf{x}) = \mathrm{tr}(\mathbf{AV}) + \boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu} \ .$$

Normality

$$\text{m.g.f.} = M_{\mathbf{x}'\mathbf{A}\mathbf{x}}^{(t)} = |\mathbf{I} - 2t\mathbf{AV}|^{-\frac{1}{2}}\exp\{-\tfrac{1}{2}\boldsymbol{\mu}'[\mathbf{I} - (\mathbf{I} - 2t\mathbf{AV})^{-1}]\mathbf{V}^{-1}\boldsymbol{\mu}\}$$

$$v(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2\mathrm{tr}(\mathbf{AV})^2 + 4\boldsymbol{\mu}'\mathbf{AVA}\boldsymbol{\mu}$$

$$K_r(\mathbf{x}'\mathbf{A}\mathbf{x}) = 2^{r-1}(r-1)![\mathrm{tr}(\mathbf{AV})^r + r\boldsymbol{\mu}'\mathbf{A}(\mathbf{VA})^{r-1}\boldsymbol{\mu}] \ .$$

Extensions: bilinear forms (normality)

$$E(\mathbf{x}_1'\mathbf{A}_{12}\mathbf{x}_2) \quad \text{and} \quad \mathrm{cov}(\mathbf{x}_1'\mathbf{A}_{12}\mathbf{x}_1,\ \mathbf{x}_3'\mathbf{A}_{34}\mathbf{x}_4) \ .$$

Searle (1971, pp. 64-66).

<u>χ^2 distribution</u> (normality)

$$\mathbf{x}'\mathbf{A}\mathbf{x} \sim \chi^{2'}[r(\mathbf{A}),\ \tfrac{1}{2}\boldsymbol{\mu}'\mathbf{A}\boldsymbol{\mu}] \quad \text{iff} \quad \mathbf{AV} \text{ is idempotent .}$$

<u>Independence of quadratic forms</u> (normality)

If $\mathbf{AVB} = \mathbf{0}$, then $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ independent

Proof: $\left.\begin{array}{l} \mathbf{A} = \mathbf{A}' \Rightarrow \mathbf{A} = \mathbf{LL}' \\ \mathbf{B} = \mathbf{B}' \Rightarrow \mathbf{B} = \mathbf{MM}' \end{array}\right\}$ for $\mathbf{L}$ and $\mathbf{M}$ of full column rank

$\therefore\ \mathbf{AVB} = \mathbf{0} \Rightarrow \mathbf{LL}'\mathbf{VMM}' = \mathbf{0}$

$\Rightarrow \mathbf{L}'\mathbf{VM} = \mathbf{0}$ because $(\mathbf{L}'\mathbf{L})^{-1}$ and $(\mathbf{M}'\mathbf{M})^{-1}$ exist.

$\therefore\ \text{cov}(\mathbf{L}'\mathbf{x}, \mathbf{x}'\mathbf{M}) = \mathbf{L}'\mathbf{VM} = \mathbf{0}$

$\Rightarrow \mathbf{L}'\mathbf{x}$ and $\mathbf{x}'\mathbf{M}$ independent

$\Rightarrow \mathbf{x}'\mathbf{A}\mathbf{x}\ (=\mathbf{x}'\mathbf{LL}'\mathbf{x})$ and $\mathbf{x}'\mathbf{B}\mathbf{x}\ (=\mathbf{x}'\mathbf{MM}'\mathbf{x})$ ind

Converse

If $\mathbf{x}'\mathbf{A}\mathbf{x}$ and $\mathbf{x}'\mathbf{B}\mathbf{x}$ independent, then $\mathbf{AVB} = \mathbf{0}$

Proof: Numerous attempts.

Many erroneous corrections.

See Searle (1971), Driscoll and Grundberg (1986), and Reid and Driscoll (1988).

LINEAR MODELS

$$E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} .$$

Examples

Regression:
$$E(\mathbf{y}) = [1 \quad \mathbf{x}_1 \quad \mathbf{x}_2]\begin{bmatrix}\mu \\ \beta_1 \\ \beta_2\end{bmatrix}$$

$$E(y_i) = \mu + x_{1i}\beta_1 + x_{2i}\beta_2 .$$

Designed experiment: randomized complete blocks, 2 treatments
3 blocks

$$\boxed{T_1 \quad T_2} \qquad \boxed{T_2 \quad T_1} \qquad \boxed{T_1 \quad T_2}$$

$$E\begin{bmatrix}y_{11}\\ y_{21}\\ y_{12}\\ y_{22}\\ y_{13}\\ y_{23}\end{bmatrix} = \begin{bmatrix}\mu + \tau_1 + b_1\\ \mu + \tau_2 + b_1\\ \mu + \tau_1 + b_2\\ \mu + \tau_2 + b_2\\ \mu + \tau_1 + b_3\\ \mu + \tau_2 + b_3\end{bmatrix} = \begin{bmatrix}1 & 1 & \cdot & 1 & \cdot & \cdot\\ 1 & \cdot & 1 & 1 & \cdot & \cdot\\ 1 & 1 & \cdot & \cdot & 1 & \cdot\\ 1 & \cdot & 1 & \cdot & 1 & \cdot\\ 1 & 1 & \cdot & \cdot & \cdot & 1\\ 1 & \cdot & 1 & \cdot & \cdot & 1\end{bmatrix}\begin{bmatrix}\mu\\ \tau_1\\ \tau_2\\ b_1\\ b_2\\ b_3\end{bmatrix} .$$

Survey data: 4, 2 and 3 people of Types 1, 2 and 3:

$$E\begin{bmatrix}y_{11}\\ y_{12}\\ y_{13}\\ y_{14}\\ y_{21}\\ y_{22}\\ y_{31}\\ y_{32}\\ y_{33}\end{bmatrix} = \begin{bmatrix}\mu + \alpha_1\\ \mu + \alpha_1\\ \mu + \alpha_1\\ \mu + \alpha_1\\ \mu + \alpha_2\\ \mu + \alpha_2\\ \mu + \alpha_3\\ \mu + \alpha_3\\ \mu + \alpha_3\end{bmatrix} = \begin{bmatrix}1 & 1 & \cdot & \cdot\\ 1 & 1 & \cdot & \cdot\\ 1 & 1 & \cdot & \cdot\\ 1 & 1 & \cdot & \cdot\\ 1 & \cdot & 1 & \cdot\\ 1 & \cdot & 1 & \cdot\\ 1 & \cdot & \cdot & 1\\ 1 & \cdot & \cdot & 1\\ 1 & \cdot & \cdot & 1\end{bmatrix}\begin{bmatrix}\mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\end{bmatrix} .$$

Estimation by least squares

Estimate $\boldsymbol{\beta}$ by the function of $\mathbf{y}$ and $\mathbf{X}$ that minimizes

$$\Sigma_t[y_t - E(y_t)]^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\frac{\partial}{\partial\boldsymbol{\beta}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \ .$$

Estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}$ that satisfies $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}$.

$(\mathbf{X}'\mathbf{X})^{-1}$ existing (e.g., regression, usually)

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \ .$$

$(\mathbf{X}'\mathbf{X})^{-1}$ not existing (most other situations)

$$\boldsymbol{\beta}^\circ = (\mathbf{X}'\mathbf{X})^{-} \mathbf{X}'\mathbf{y} \ .$$

Variances

$$\text{With } E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}, \text{ define } \mathbf{e} = \mathbf{y} - E(\mathbf{y}) \ .$$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$$

Attribute $\sigma^2\mathbf{I}$ to $\text{var}(\mathbf{e}) = \text{var}(\mathbf{y})$

$$\text{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

$$\text{var}(\boldsymbol{\beta}^\circ) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\sigma^2 \ .$$

Many values of $\boldsymbol{\beta}^{\circ}$

Solutions to $\mathbf{X'X}\boldsymbol{\beta}^{\circ} = \mathbf{X'y}$;

With z arbitrary, $\boldsymbol{\beta}^{\circ} = (\mathbf{X'X})^{-}\mathbf{X'y} + [\mathbf{I} - (\mathbf{X'X})^{-}\mathbf{X'X}\,]\mathbf{z}$.

Properties of $(\mathbf{X'X})^{-}$

Definition: $\mathbf{X'X}(\mathbf{X'X})^{-}\mathbf{X'X} = \mathbf{X'X}$

Transpose: $\mathbf{X'X}(\mathbf{X'X})^{-'}\mathbf{X'X} = \mathbf{X'X} \Rightarrow (\mathbf{X'X})^{-'}$ is a $(\mathbf{X'X})^{-}$.

For real matrices, $\mathbf{PX'X} = \mathbf{QX'X} \Rightarrow \mathbf{PX'} = \mathbf{QX'}$

$\mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'X} = \mathbf{X}$.

Invariance: $\mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'}$ is invariant to $(\mathbf{X'X})^{-}$.

Symmetry: $\mathbf{X}(\mathbf{X'X})^{-}\mathbf{X'}$ is symmetric.

$(\mathbf{X'X})^{\sim} = (\mathbf{X'X})^{-}\mathbf{X'X}(\mathbf{X'X})^{-'}$ is symmetric

and $(\mathbf{X'X})^{\sim}\mathbf{X'X}(\mathbf{X'X})^{\sim} = (\mathbf{X'X})^{\sim}$.

Useful functions of $\boldsymbol{\beta}^\circ$

$\mathbf{q}'\boldsymbol{\beta}^\circ$ is invariant to $\boldsymbol{\beta}^\circ$

$$\text{iff} \quad \mathbf{q}' = \mathbf{t}'\mathbf{X} \text{ for some } \mathbf{t}$$

$$\equiv \quad \text{iff} \quad \mathbf{q}' = \mathbf{q}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X} \ .$$

Then

$$E(\mathbf{q}'\boldsymbol{\beta}^\circ) = \mathbf{q}'\boldsymbol{\beta}$$

$\mathbf{q}'\boldsymbol{\beta}$ is said to be "estimable function"

$$\widehat{\mathbf{q}'\boldsymbol{\beta}} = \mathbf{q}'\boldsymbol{\beta}^\circ \ ,$$

i.e.,

$$\text{B.L.U.E.}(\mathbf{q}'\boldsymbol{\beta}) = \mathbf{q}'\boldsymbol{\beta}^\circ$$

with

$$v(\mathbf{q}'\boldsymbol{\beta}^\circ) = \mathbf{q}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{q}\sigma^2, \text{ invariant to } (\mathbf{X}'\mathbf{X})^{-}$$

$$\leq v(\text{any other linear, unbiased estimator of } \mathbf{q}'\boldsymbol{\beta}) \ .$$

In practice "estimable" functions are often (but not always) useful.

<u>Sums of squares</u>

Residual: after fitting model

$$\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}^\circ = \mathbf{X}(\mathbf{X}'\mathbf{X})\ \mathbf{X}'\mathbf{y}, \text{ invariant to } \boldsymbol{\beta}^\circ$$

$$\text{SSE} = (\mathbf{y} - \hat{\mathbf{y}})'(\mathbf{y} - \hat{\mathbf{y}}) = \mathbf{y}'\mathbf{y} - \boldsymbol{\beta}^{\circ\prime}\mathbf{X}'\mathbf{y}$$

$$E(\text{SSE}) = (N - r_{\mathbf{X}})\sigma^2 \ .$$

<u>Testing linear hypotheses</u>

$$H : \mathbf{K}'\boldsymbol{\beta} = \mathbf{m} \qquad \text{for } \mathbf{K}'\boldsymbol{\beta} \text{ estimable}$$

$$\mathbf{K}' \text{ of full row rank}$$

$$r_{\mathbf{K}'} \le r_{\mathbf{X}}$$

$$Q = (\mathbf{K}'\boldsymbol{\beta}^\circ - \mathbf{m})'[\mathbf{K}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{K}]^{-1}(\mathbf{K}'\boldsymbol{\beta}^\circ - \mathbf{m})$$

$$F = \frac{Q/r_{\mathbf{K}'}}{\text{SSE}/(N - r_{\mathbf{X}})} \quad \text{test } H \ .$$

Degrees of freedom: $r_{\mathbf{K}'}$ and $(N - r_{\mathbf{X}})$.

Variance structure other than $\sigma^2 \mathbf{I}$

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$$\mathbf{u} = \left\{_c \; \mathbf{u}_i\right\}_{i=1}^{t} \qquad \mathbf{Z} = \left\{_r \; \mathbf{Z}_i\right\}_{i=1}^{t}$$

$$\mathbf{u}_0 = \mathbf{e} \qquad \mathbf{Z}_0 = \mathbf{I}$$

$$\dot{\mathbf{u}} = \begin{bmatrix} \mathbf{u}_0 \\ \mathbf{u} \end{bmatrix} \qquad \dot{\mathbf{Z}} = [\mathbf{Z}_0 \quad \mathbf{Z}]$$

$$\mathbf{D} = \mathrm{var}(\mathbf{u}) = \left\{_d \; \sigma_i^2 \mathbf{I}_{n_i}\right\}_{i=1}^{t}; \quad \dot{\mathbf{D}} = \begin{bmatrix} \sigma_0^2 \mathbf{R}_N & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}$$

$$\mathbf{V} = \mathrm{var}(\mathbf{y}) = \dot{\mathbf{Z}}\dot{\mathbf{D}}\dot{\mathbf{Z}}$$

$$= \mathbf{Z}\mathbf{D}\mathbf{Z}' + \sigma_0^2 \mathbf{R}_N \; .$$

Fourth moments (for $\mathbf{R} = \mathbf{I}$)

$$\mathbf{F} = \mathrm{var}[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) * (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] = \mathrm{var}[\mathrm{vec}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})']$$

$$= (\mathbf{V} * \mathbf{V})(\mathbf{I}_{N^2 \times N^2} + \mathbf{I}_{N,N})$$

$$+ (\dot{\mathbf{Z}} * \dot{\mathbf{Z}})\left\{_d \; \mathrm{vec}\left\{_d \; \gamma_i \sigma_i^4 \mathbf{I}_{n_i}\right\}_{i=0}^{t}\right\}(\dot{\mathbf{Z}}' * \dot{\mathbf{Z}}')$$

$$= (\mathbf{V} * \mathbf{V})(\mathbf{I}_{N^2 \times N^2} + \mathbf{I}_{N,N}), \text{ under normality.}$$

Estimating $\boldsymbol{\beta}$ and predicting $\mathbf{u}$

$$\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta}^\circ = \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} .$$

Also

$$\begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{D}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}^\circ \\ \tilde{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{bmatrix} ; \quad \text{same } \boldsymbol{\beta}^\circ .$$

And

$$E(\mathbf{u}|y) = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

$$\tilde{\mathbf{u}} = \mathbf{D}\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}^\circ) : \qquad \text{B.L.U.P.}(\mathbf{u}) .$$

In the 1-way classification

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad j = 1,\cdots,n$$

$$\tilde{u}_i = \frac{nh}{1 + (n-1)h}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) \qquad \text{for } h = \sigma^2_\alpha/(\sigma^2_\alpha + \sigma^2_e)$$

MULTIVARIATE STATISTICS

p random variables : 1, 2, $\cdots$, p .

n observations on each: $\mathbf{Y}_{n\times p} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ \vdots & & & \vdots \\ y_{n1} & y_{n2} & & y_{np} \end{bmatrix}$.

Each row of $\mathbf{Y}$: $\mathbf{y}_i' \sim N(\boldsymbol{\mu}', \mathbf{V})$

Rows uncorrelated (and independent, from normality) .

<u>Maximum Likelihood Estimation</u>

For i'th row: $L_i = \dfrac{\exp - \frac{1}{2}(\mathbf{y}_i' - \boldsymbol{\mu}')\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})}{(2\pi)^{\frac{1}{2}p}|\mathbf{V}|^{\frac{1}{2}}}$, $\qquad i = i,\cdots,n$.

For all rows: $L = \pi L_i$ because rows independent

$$= \frac{\exp - \frac{1}{2}\Sigma_i(\mathbf{y}_i' - \boldsymbol{\mu}')\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu})}{(2\pi)^{\frac{1}{2}np}|\mathbf{V}|^{\frac{1}{2}}} .$$

$$2\log L = \Sigma_i(\mathbf{y}_i' - \boldsymbol{\mu}')\mathbf{V}^{-1}(\mathbf{y}_i - \boldsymbol{\mu}) - np\log 2\pi + n\log|\mathbf{V}^{-1}| .$$

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = 0 \quad \Rightarrow \quad \Sigma_i \mathbf{V}^{-1}\hat{\boldsymbol{\mu}}' = \Sigma_i \mathbf{V}^{-1}\mathbf{y}_i' \quad \Rightarrow \quad \hat{\boldsymbol{\mu}} = \Sigma_i \mathbf{y}_i/n = \bar{\mathbf{y}}.$$

Rewrite

$$2\log L = -\mathrm{tr}[(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}')\mathbf{V}^{-1}(\mathbf{Y}' - \boldsymbol{\mu}\mathbf{1}')] - np\log 2\pi + n\log|\mathbf{V}^{-1}|$$

$$= -\mathrm{tr}[(\mathbf{Y}' - \boldsymbol{\mu}\mathbf{1}')(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu}')\mathbf{V}^{-1}] - np\log 2\pi + n\log|\mathbf{V}^{-1}| .$$

Two pertinent results

Henderson and Searle, 1979, *Canadian J. of Statistics* **7**, 65-81.

$$\left.\begin{aligned}\partial[\mathrm{tr}(\mathbf{AX})]/\partial\mathbf{X} &= \mathbf{A} + \mathbf{A}' - \mathrm{diag}\ \mathbf{A}\\ \partial[\log|\mathbf{X}|]/\partial\mathbf{X} &= 2\mathbf{X}^{-1} - \mathrm{diag}\ \mathbf{X}^{-1}\end{aligned}\right\}\ \text{for } \mathbf{X} = \mathbf{X}'\ .$$

where diag **A** is the diagonal matrix of the diagonal elements of **A**.

Application to 2logL

$$\frac{\partial}{\partial\mathbf{V}^{-1}}2\log L = 0 \quad \Rightarrow \quad -(2\hat{\mathbf{M}} - \mathrm{diag}\ \hat{\mathbf{M}}) + n(2\hat{\mathbf{V}} - \mathrm{diag}\ \hat{\mathbf{V}}) = \mathbf{0}$$

$$\text{for } \mathbf{M} = (\mathbf{Y}' - \boldsymbol{\mu}\mathbf{1}')(\mathbf{Y} - \mathbf{1}\boldsymbol{\mu})$$

$$\Rightarrow \quad 2(n\hat{\mathbf{V}} - \hat{\mathbf{M}}) - \mathrm{diag}(n\hat{\mathbf{V}} - \hat{\mathbf{M}}) = \mathbf{0}$$

$$\Rightarrow \quad \hat{\mathbf{V}} = \hat{\mathbf{M}}/n \quad \text{since } \mathbf{V} \text{ and } \mathbf{M} \text{ are symmetric .}$$

$$\hat{\mathbf{V}} = (\mathbf{Y}' - \bar{\mathbf{y}}_{.}\mathbf{1}')(\mathbf{Y} - \mathbf{1}\bar{\mathbf{y}}_{.}')\ .$$

Wishart matrix:

$$\mathbf{Y}'\mathbf{Y} = \text{when } \boldsymbol{\mu} = \mathbf{0}$$

$$E(\mathbf{Y}'\mathbf{Y}) = E\sum_{i=1}^{n}\mathbf{y}_i\mathbf{y}_i' = n\mathbf{V}$$

$$\begin{aligned}\mathrm{var}[\mathrm{vec}(\mathbf{Y}'\mathbf{Y})] &= \Sigma_i \mathrm{var}[\mathrm{vec}(\mathbf{y}_i\mathbf{y}_i')]\\ &= n(\mathbf{V} * \mathbf{V})(\mathbf{I}_{n^2\times n^2} + \mathbf{I}_{n,n})\ .\end{aligned}$$

Linear transformations

$(\mathbf{Y}')_{p\times n}$ has uncorrelated columns $\sim (\mathbf{m}_j, \mathbf{V})$

$\mathbf{BY}'$ " " " $\sim (\mathbf{Bm}_j, \mathbf{BVB}')$

$\text{vec}(\mathbf{BY}') \sim [\text{vec}(\mathbf{BM}), \mathbf{I} * \mathbf{BVB}']$

$E(\mathbf{Y}'\mathbf{C}) = \mathbf{MC}$.

$$\begin{aligned}
\text{var}[\text{vec}(\mathbf{Y}'\mathbf{C})] &= \text{var}[(\mathbf{C}' * \mathbf{I})\text{vec } \mathbf{Y}'] \\
&= (\mathbf{C}' * \mathbf{I})\text{var}(\text{vec } \mathbf{Y}')(\mathbf{C}' * \mathbf{I})' \\
&= (\mathbf{C}' * \mathbf{I})(\mathbf{I} * \mathbf{V})(\mathbf{C} * \mathbf{I}) \\
&= \mathbf{C}'\mathbf{C} * \mathbf{V} .
\end{aligned}$$

$$\begin{aligned}
\text{var}[\text{vec}(\mathbf{YA})] &= \text{var}[(\mathbf{A}' * \mathbf{I})\text{vec } \mathbf{Y}] \\
&= \text{var}[(\mathbf{A}' * \mathbf{I})\mathbf{I}_{n,p}\text{vec}\mathbf{Y}'] \\
&= (\mathbf{A}' * \mathbf{I})\mathbf{I}_{n,p}(\mathbf{I} * \mathbf{V})\mathbf{I}_{p,n}(\mathbf{A}' * \mathbf{I})' \\
&= (\mathbf{A}' * \mathbf{I})(\mathbf{V} * \mathbf{I})(\mathbf{A} * \mathbf{I}) \\
&= \mathbf{A}'\mathbf{VA} * \mathbf{I}
\end{aligned}$$

A TASTE OF FOUR (FROM MANY) OTHER TOPICS

Multivariate linear model

$$\mathbf{Y}_{n\times p} = \mathbf{X}_{n\times q}\mathbf{B}_{q\times p} + \mathbf{E}_{n\times p} \; .$$

Reduce to univariate

$$\mathrm{vec}\mathbf{Y} \sim [(\mathbf{I}_p * \mathbf{X})\mathrm{vec}\mathbf{B},\ \mathbf{\Sigma} * \mathbf{I}_p]$$

where the n rows of E are n.i.i.d.$(\mathbf{0}, \mathbf{\Sigma})$.

$$F = \frac{Q/r_{\mathbf{K}'}}{SSE/(N - r_{\mathbf{X}})} \text{ of univariate}$$

reduces to Hotelling's T_0^2-statistic for testing that some rows of $\mathbf{B}$ have pre-assigned values.

Finite Markov Chains

$\mathbf{P} = \{p_{ij}\}$ with $p_{ij} = \Pr(\text{state } i \text{ to state } j)$.

$\mathbf{P1} = \mathbf{1}$.

Solve $(\mathbf{I} - \mathbf{P})\mathbf{u} = \mathbf{c}$; $(\mathbf{I} - \mathbf{P})\mathbf{1} = \mathbf{0}$.

$$\mathbf{u} = (\mathbf{I} - \mathbf{P})^{-}\mathbf{c} \ .$$

Theorem For $\mathbf{A1} = \mathbf{0}$, a value of $\mathbf{A}^{-}$ is $(\mathbf{A} + \mathbf{xy}')^{-1}$,

for $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{y}'\mathbf{1} = 1$;

and then (i) $\mathbf{A}(\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{x} = \mathbf{0}$

(ii) $\mathbf{y}'(\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{x} = 1$

(iii) $(\mathbf{A} + \mathbf{xy}')^{-1}$ is a value of $(\mathbf{xy}')^{-}$.

Proof $\mathbf{y}'\mathbf{1} = 1 \Rightarrow (\mathbf{A} + \mathbf{xy}')\mathbf{1} = \mathbf{x}$

$$\therefore \quad (\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{x} = \mathbf{1}$$

$$\therefore \quad \mathbf{A}(\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{x} = \mathbf{A1} = \mathbf{0} \qquad \text{(i)}$$

$$\therefore \quad \mathbf{A}(\mathbf{A} + \mathbf{xy}')^{-1}(\mathbf{A} + \mathbf{xy}) = \mathbf{A}$$

becomes $\mathbf{A}(\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{A} = \mathbf{A}$.

And $(\mathbf{A} + \mathbf{xy}')(\mathbf{A} + \mathbf{xy}')^{-1}\mathbf{x} = \mathbf{x}$

becomes $\mathbf{xy}'(\mathbf{A} + \mathbf{xy}')\mathbf{x} = \mathbf{x}$.

But $\mathbf{x}'\mathbf{x} \neq 0 \Rightarrow \mathbf{y}'(\mathbf{A} + \mathbf{xy}')\mathbf{x} = 1$. (ii)

$$\Rightarrow \quad \mathbf{xy}'(\mathbf{A} + \mathbf{xy}')\mathbf{xy}' = \mathbf{xy}' \ . \qquad \text{(iii)}$$

<u>Fourth moments</u>

(Neudecker and Wansbeek, 1983, *Canadian Journal of Statistics* **11**, 221-231.)

$\mathbf{x}_{p\times 1} \sim (\boldsymbol{\mu}_1, \mathbf{V}_1 + \boldsymbol{\mu}_1\boldsymbol{\mu}_1')$ independent of $\mathbf{y}_{q\times 1} \sim (\boldsymbol{\mu}_2, \mathbf{V}_2 + \boldsymbol{\mu}_2\boldsymbol{\mu}_2)$.

$$E(\mathbf{xy}' * \mathbf{xy}') = \mathrm{vec}\mathbf{V}_1(\mathrm{vec}\mathbf{V}_2)'$$

$$E(\mathbf{xy}' * \mathbf{yx}') = (\mathbf{V}_1 * \mathbf{V}_2)\mathbf{I}_{q,p} = \mathbf{I}_{q,p}(\mathbf{V}_2 * \mathbf{V}_1)$$

$$E(\mathbf{x} * \mathbf{x} * \mathbf{y} * \mathbf{y}) = \mathrm{vec}\mathbf{V}_1 * \mathrm{vec}\mathbf{V}_2$$

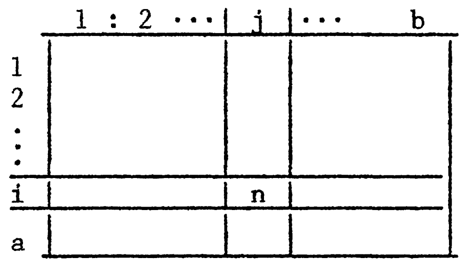
$$E(\mathbf{x} * \mathbf{y} * \mathbf{x} * \mathbf{y}) = (\mathbf{I}_p * \mathbf{I}_{p,q} * \mathbf{I}_q)(\mathrm{vec}\mathbf{V}_1 * \mathrm{vec}\mathbf{V}_2)$$

$$E(\mathbf{x} * \mathbf{y} * \mathbf{y} * \mathbf{x}) = (\mathbf{I}_p * \mathbf{I}_{p,qq})(\mathrm{vec}\mathbf{V}_1 * \mathrm{vec}\mathbf{V}_2)$$

··· and so on.

<u>Variance components models</u>

2-way classification



$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

$$\mathbf{y} = \left\{\left\{\left\{y_{ijk}\right\}_{k=1}^{n}\right\}_{j=1}^{b}\right\}_{i=1}^{a} \text{, lexicon order}$$

$$\begin{aligned}\mathbf{y} &= (\mathbf{1}_a * \mathbf{1}_b * \mathbf{1}_n)\mu + (\mathbf{I}_a * \mathbf{1}_b * \mathbf{1}_n)\boldsymbol{\alpha} \\ &\quad + (\mathbf{1}_a * \mathbf{I}_b * \mathbf{1}_n)\boldsymbol{\beta} + (\mathbf{I}_a * \mathbf{I}_b * \mathbf{1}_n)\boldsymbol{\gamma} + (\mathbf{I}_a * \mathbf{I}_b * \mathbf{I}_n)\mathbf{e}\end{aligned}$$

$$\begin{aligned}\mathbf{V} = \mathrm{var}(\mathbf{y}) &= (\mathbf{I}_a * \mathbf{J}_b * \mathbf{J}_n)\sigma^2_{\boldsymbol{\alpha}} + (\mathbf{J}_a * \mathbf{I}_b * \mathbf{J}_n)\sigma^2_{\boldsymbol{\beta}} \\ &\quad + (\mathbf{I}_a * \mathbf{I}_b * \mathbf{J}_n)\sigma^2_{\gamma} + (\mathbf{I}_a * \mathbf{I}_b * \mathbf{I}_n)\sigma^2_{e} \\ &= \Sigma_t \theta_t \mathbf{K}_t : \ \mathbf{K}\text{s commutative .}\end{aligned}$$